Proceedings of the 33rd
Conference on Decision and Control
Lake Buena Vista, FL - December 1994

Observability/Identifiability of Rigid Motion under Perspective Projection

Stefano Soatto

California Institute of Technology 116-81–Pasadena, CA 91125 · soatto@systems.caltech.edu

Abstract

The "visual motion estimation" problem concerns the reconstruction of the motion of an object viewed under projection. This paper addresses the feasibility of such a problem when the object is represented as a "rigid" set of point-features in the Euclidean 3D space. We represent rigid motion as a point on the so-called "essential manifold" and show that it is globally observable from perspective projections under some general position conditions. Such conditions hold when the path of the viewer and the visible objects cannot be embedded in a quadric surface of $\mathbf{R}^3$.

1. Introduction

Animals require the ability to estimate the relative motion between themselves and the environment when facing everyday tasks such as walking, avoiding obstacles, grasping objects. Only recently, however, have dynamic estimation and control techniques given encouraging results for designing automatic systems which mimic such abilities [8, 19, 15]. If we restrict our attention to motions inside a "static scene", the rigid motion constraint and the perspective projection map *define* a nonlinear dynamical model. Motion estimation may be formalized in terms of parameter identification and/or state estimation of such a model. Traditionally, the estimation task has been performed using Extended Kalman Filters (EKF) [6, 17, 18]. A crucial issue in dynamic estimation/identification is the observability of the model, or the identifiability of its parameters. We will see that the model which "defines" the visual motion problem for feature points in the Euclidean 3D space is neither linearly observable nor locally weakly observable. It is possible, as we will see, to reduce the set of locally indistinguishable states by imposing metric constraints on the state space; however, the model suffers some structural limitations which make the local-linearization based methods poorly conditioned and not robust enough to be used in real world applications. Rigid motion is indeed globally observable from perspective projections, once the problem is formulated in the appropriate topological space. In this paper we analyze a novel formulation for motion estimation [29] in terms of *identification of a nonlinear implicit model with the parameters living on a manifold*, called the "essential manifold". Using results from the computational vision literature [10, 23], we show that this model is globally observable without any Lie differentiation under general position conditions. Such (sufficient) conditions are met when the object and the path of the center of projection cannot be embedded in a quadric surface, and may be verified using a simple rank test. The use of dynamic observers to estimate scene structure and/or motion dates back to the eighties [4, 11, 14, 22]. Many current schemes (for instance [1, 22, 27]) are based upon minor variations of the same model, and none of them addresses the issue of its observability. Our work is somehow complementary to [7, 12], in which the feasibility of structure estimation is assessed. We study instead the problem of motion estimation for unknown structure. Once motion is known, structure is linearly observable from the rigid motion model.

2. Visual motion estimation: statement and formalization of the problem

Let us now consider a simple paradigm, in which $p_i \in \mathbf{R}^3$ is a salient point in the scene; $\mathbf{X}_i \doteq [X, Y, Z]_i^T$ are its coordinates with respect to an orthonormal reference frame centered in the pupil of the viewer, with the Z axis pointing forward and X, Y arranged as to form a right-handed frame. Let $[V \ , \ \Omega]^T \in se(3)$ represent the canonical (exponential) coordinates of the rigid motion of the viewer [25]. As the viewer moves, each point describes a vector field f on $\mathbf{R}^3$; in the viewer-centered representation we have

$$f(\mathbf{X}, V(t), \Omega(t)) = \Omega(t) \wedge \mathbf{X} + V(t).$$

If we consider motion between two time instants t and $t + \tau$, and the velocity is held constant between the two samples, we have $f(\mathbf{X}, T(t,\tau), R(t,\tau)) = R(t,\tau)\mathbf{X} + T(t,\tau)$ where (T, R) are related to (V, Ω) via [25]

$$R(t,\tau) \doteq e^{(\Omega(t)\wedge)\tau} \qquad (1)$$

$$T(t,\tau) \doteq \mathcal{T}(\tau,\Omega(t))V(t)$$
$$\mathcal{T}(\tau,\Omega) \doteq \frac{1}{\|\Omega\|}\left[\left(I - e^{(\Omega\wedge)\tau}\right)\Omega\wedge + \Omega\Omega^T\tau\right]. \quad (2)$$

In the following we assume a constant sampling rate $\tau = 1$. We measure the perspective projection π up to some noise. The map π is the trivial association of each $p \mid p \neq 0$ with its projective coordinates as an element of $\mathbb{RP}^2$: if $\mathbf{X} = [X\ Y\ Z]^T$ are the euclidean coordinates of p, then we denote with $\mathbf{x} \doteq [x\ y\ 1]^T = [\ \frac{X}{Z}\ \ \frac{Y}{Z}\ \ 1\ ]^T$ its projective coordinates. In summary, when we represent the scene structure using points in the Euclidean 3D space, the visual motion problem is *defined* by the constraints of rigid motion and perspective projection. In the viewer-centered instantaneous representation we have

$$\begin{cases} \dot{\mathbf{X}}_i = \Omega \wedge \mathbf{X}_i + V & \mathbf{X}_i(0) = \mathbf{X}_{i0} \\ \mathbf{x}_i = \pi(\mathbf{X}_i) + \nu_i & \nu_i \in \mathcal{N}(0, R_i) \end{cases} \quad \forall i = 1:n \quad (3)$$

where ν_i stands for an error in measuring the coordinates of the projection of the point p_i. Solving the visual motion problem consists of estimating $\mathbf{X}_i, V$ and Ω for all the visible points p_i, i.e. reconstructing both the input and the initial state of the above system from its noisy output. Alternatively, motion may be viewed as a vector of unknown parameters in the model (3), which have to be identified.

3. Visual motion estimation as a filtering problem

Motion estimation may be viewed as an inversion problem for the model (3) when the initial state (structure) is unknown. It is well known that under certain conditions on the relative degree, it is possible to invert a nonlinear system [16]. In order to do that, we compute Lie derivatives of the output along the state vector fields until the components of the input appear. If the coupling matrix is nonsingular, we may invert it and reconstruct the input of the system from bracket combinations of its output. In our case the model is *driftless* and, therefore, all the components of the input appear at the first level of differentiation. The first time derivative of the output is in fact $\dot{\mathbf{x}}_i(t) = \mathcal{C}_{.i}\ [\ V(t)\ \ \Omega(t)\ ]^T$ where $\mathcal{C}_{.i} \doteq \begin{bmatrix} 1 & 0 & -x_i & -x_iy_i & 1+x_i^2 & -y_i \\ 0 & 1 & -y_i & -1-y_i^2 & x_iy_i & x_i \end{bmatrix}$. Once we observe enough points, we have an overdetermined system which we may solve for the motion parameters in a least-squares fashion: $[\ \hat{V}(t)\ \ \hat{\Omega}(t)\ ]^T \doteq \mathcal{C}^{\dagger}\dot{\mathbf{x}}$ where the symbol $\dagger$ denotes the pseudo-inverse. Note that $\mathcal{C}_i$ depends on the image measurements $\mathbf{x}_i$, and on the depth of each point, Z_i, which we do not know. In order to reconstruct the initial depth it is necessary to *observe* it. Dynamic observers are, in essence, computing differentiations of the output until the matrix which couples the initial condition and the derivatives of the output (observability matrix or observability codistribution) has full rank. In our case, however, *both the input and the initial state appear at the same level of differentiation*, since $L_f^k\pi(\mathbf{X}) = 0\ \forall\ k > 0$. Therefore, we can hope to recover either motion or depth with this technique, but not both. See [13, 30] for more details on this formulation.

Because the model described above has no drift dynamics, left-inversion/state-estimation reduces to a static (instantaneous) procedure and hence it does not exploit the noise rejection properties of dynamic observers. One possible way to proceed, based on the above considerations, is to use "dynamic extension". Instead of considering motion as the input of the system, we consider as input its time derivative, and insert motion into the state dynamics. We arrive to the augmented model:

$$\begin{cases} \dot{\mathbf{X}}_i = \Omega \wedge \mathbf{X}_i + V & \mathbf{X}_i(0) = \mathbf{X}_{i0} \\ \dot{V} = f_V(V, \nu_V) & V(0) = V_0 \\ \dot{\Omega} = f_\Omega(\Omega, \nu_\Omega) & \Omega(0) = \Omega_0 \\ \mathbf{x}_i = \pi(\mathbf{X}_i) + \nu_i & \end{cases} \quad \forall i = 1:n. \quad (4)$$

We denote with $\bar{f}$ the augmented state vector field of the above model. Since f_V and f_Ω are unknown, the visual motion problem may be formulated as an "unknown-input/state estimation" problem. However, one may want to exploit some a priori information about f_V, f_Ω, for example a simplified dynamical model when the camera is mounted onto a moving vehicle. In absence of such information, f_V and f_Ω may describe a statistical model. The simplest case is $f_V = f_\Omega \cong 0$, which corresponds to constant velocity (or "small acceleration"). A model often used is Brownian motion. A crucial issue in state estimation using observers is, of course, the *observability* of the model, which is addressed in the following section.

4. Perspective local observability of rigid motion

In this section we study the local observability of the model (3). In the case of constant velocity (or small acceleration), the model is not locally observable. However, by enforcing metric constraints on the state space, it is possible to reduce the set of locally indistinguishable states. Some definitions and standard results on local observability may be found in [16].

4.1. Linear observability

Consider the linearization of the model (4): define $A \doteq \frac{\partial \bar{f}(x)}{\partial x}, C \doteq \frac{\partial \pi(x)}{\partial x}$. Suppose for simplicity $n = 1$:

$$CA^i = \frac{1}{Z}\mathcal{A}_i\ [\ (\Omega\wedge)^i\ \ \ (\Omega\wedge)^{i-1}\ \ \ -(\Omega\wedge)^{i-1}(\mathbf{X}\wedge)\ ]$$

where $\mathcal{A}_i \doteq \begin{bmatrix} 1 & 0 & -x_i \\ 0 & 1 & -y_i \end{bmatrix}$. The observability matrix for the linearized system has rank 5, in face of a state space of dimension 9. The linearized system is, therefore, not observable, and we say that the original model is not linearly observable.

4.2. Local observability

The local observability space $\mathcal{O}$ is defined as the set of the output functions and all their possible Lie derivatives along vector fields in the accessibility algebra [16]. Under small acceleration, the vector field in (3) is autonomous and, therefore, the observability space is spanned by $\{\pi, L_{\bar{f}}\pi, \ldots, L_{\bar{f}}^k\pi \ldots\}$, where $\bar{f}$ is the state vector field. The observability codistribution is spanned by $d\mathcal{O} \doteq \{dh \mid h \in \mathcal{O}\}$. The state manifold is $\mathbf{R}^9$, intended as a local coordinatization of $SE(3) \times \mathbf{R}^3$. The rank of the observability codistribution reaches its maximum of 8 after three levels of Lie differentiation, after which it becomes involutive. The null space of the observability codistribution, in case of non-zero forward translation, is:

$$\text{Span}\left[\begin{array}{ccccccccc} \frac{X}{V_Z} & \frac{Y}{V_Z} & \frac{Z}{V_Z} & \frac{V_X}{V_Z} & \frac{V_Y}{V_Z} & 1 & 0 & 0 & 0 \end{array}\right] \quad (5)$$

and similarly in the case of zero forward translation but non-zero lateral translation. The set of states which are indistinguishable from $[\mathbf{X}_0\ V_0\ \Omega_0]^T$ is

$$I\left(\left[\begin{array}{c} \mathbf{X}_0 \\ V_0 \\ \Omega_0 \end{array}\right]\right) = \left\{\left[\begin{array}{c} \mathbf{X}_0\frac{V_{0_3}+s}{V_{0_3}} \\ V_0\frac{V_{0_3}+s}{V_{0_3}} \\ \Omega_0 \end{array}\right] \mid s \in \mathbf{R}\right\},$$

for $V_{0_3} \neq 0$ and similarly for the other cases. For pure rotation, a basis of the null space of the observability codistribution is $\left[\begin{array}{ccccccccc} \frac{X}{Z} & \frac{Y}{Z} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$, and all the points with the same projective coordinates are indistinguishable.

4.3. Global scale ambiguity: metric constraints on the state manifold

Consider the solution $\mathbf{X}_i(t, \mathbf{X}_{i0}, V_0, \Omega_0)$ of (4) starting from the initial conditions $\mathbf{X}_{i0}, V_0, \Omega_0$ when the motion is held constant $V(t) = V_0\,; \Omega(t) = \Omega_0$:

$$\mathbf{X}_i(t) = \begin{cases} e^{(\Omega_0\wedge)t}\mathbf{X}_{i0} + \mathcal{T}(t, \Omega_0)V_0 & \text{if } \|\Omega_0\| \neq 0 \\ V_0 t & \text{otherwise} \end{cases}$$

where $\mathcal{T}$ is defined in eq. (2). It is easily seen that

$$\mathbf{X}_i(t, \alpha\mathbf{X}_{i0}, \alpha V_0, \Omega_0) = \alpha\mathbf{X}_i(t, \mathbf{X}_{i0}, V_0, \Omega_0)$$

for all $\mathbf{X}_{i0}, V_0, \Omega_0, t, \alpha$. Since for perspective projection we have $\mathbf{x}_i(t) = \pi(\mathbf{X}_i) = \pi(\alpha\mathbf{X}_i)$, we conclude that any initial condition $\alpha_1\mathbf{X}_{i0}, \alpha_1 V_0, \Omega_0$ is indistinguishable from $\alpha_2\mathbf{X}_{i0}, \alpha_2 V_0, \Omega_0$, for any possible $\alpha_1, \alpha_2 \in \mathbf{R}$.

This one-dimensional unobservable space is very familiar, as we experience that an object translating in front of us produces the same impression of a similar one which is "twice as big, twice as far, and moving twice as fast". However, we may impose norm constraints upon the visible objects or upon the translational velocity in order to get rid of the scale factor ambiguity. For example, if we impose $\|V_0\| = 1$, two initial conditions are indistinguishable only if $\alpha_1 = \pm\alpha_2$. There are still some aspects of the model (4) which have not been elucidated: we know that, if an object is visible, it must be in front of the observer, i.e. $Z_i \geq 0\ \forall\ i$. Moreover, no points are allowed to lie on the focal plane $Z = 0$ (plane at infinity), and therefore $\alpha_1 = \alpha_2$. If we apply such metric constraints to the locally unobservable codistribution, we can reduce the set of indistinguishable states to the trivial set. However, an appropriate model should include such constraints *explicitly* into the state manifold. This may be done at the price of transforming the state from the linear space $\mathbf{R}^9$ to the differentiable manifold with boundary $\mathbf{R}^2 \times H^1 \times \mathrm{S}^2 \times \mathbf{R}^3$ (H^1 is the half space of dimension one, and S^2 is the two-sphere [3]). We now summarize some of the limitations of the model (4):

- The model is not locally observable. Metric constraints which makes the model observable are not explicitly encoded in the state representation.
- Three levels of Lie bracketing are needed to cover the observable part of the state space. We know it is possible to estimate motion and structure from the first derivative of the projection of the points (optical flow) [10, 20].
- The model has the property of being "block diagonal" with respect to the structure, so that the states corresponding to different points are independent. Therefore, adding more points does not improve the estimate of motion. Indeed, that *the more points are visible, the better the perception of motion ought to be* proves to be highly intuitive.

5. Global observability: motion estimation as identification of an implicit dynamical model

In this section we describe an alternative formalization of the visual motion problem which has been presented in [29]. It is based upon a motion representation first introduced by Longuet-Higgins [20]. Motion estimation is viewed as the problem of identifying a system in exterior differential form [5] with parameters on a manifold, called the "essential manifold" [29]. We show that the model is globally observable/identifiable with zero level of differentiation for any number of visible points. When more points are available, the redundancy may be exploited in order to reduce the effect of the measurement noise.

5.1. The "essential model"

Consider a point in 3D space, with coordinates $\mathbf{X}_i(t)$ in the viewer's reference. Let $\mathbf{X}_i(t+\tau)$ be the coordinates after a rigid motion of the viewer (T, R), of which (V, Ω) are the canonical coordinates [25] as in equations (1)-(2)[1]. It is immediate to see that $\mathbf{X}_i(t), \mathbf{X}_i(t+\tau)$ and T are coplanar,

[1] Note that T in this section differs from the one defined in the previous section. Rigid motion is represented here as $\mathbf{X}(t+\tau) = R(\mathbf{X}(t) - T)$, for consistency with the notation of [20].

and hence their triple product is zero. Once expressed in a common reference, for example the viewer's at time t, the coplanarity constraint becomes [20]

$$\mathbf{X}_i^T(t+\tau)R(T \wedge \mathbf{X}_i(t)) = 0 \qquad \forall\, i = 1:n.$$

The same relationship holds for $\mathbf{x}_i(t)$ and $\mathbf{x}_i(t+\tau)$, since they represent the projective coordinates of $\mathbf{X}_i(t)$ and $\mathbf{X}_i(t+\tau)$; $T\wedge \in so(3)$ is a skew symmetric matrix. After defining the *essential matrix* [20] as $\mathbf{Q} \doteq R(T\wedge)$, the *essential constraint* is

$$\mathbf{x}_i^T(t+\tau)\mathbf{Q}\mathbf{x}_i(t) = 0 \qquad \forall\, i = 1:n. \quad (6)$$

Since there is an arbitrary scale factor in the above equality, we impose $\|\mathbf{Q}\|_2 = \|T\| = 1$. The essential matrix was first introduced by Longuet-Higgins [20], together with a quasi-linear batch technique for estimating structure and motion from two views and more than 8 visible points. His technique was then extended and developed in [10, 29, 32]. The essential matrices are points of the space

$$\tilde{E} \doteq \{RS | R \in SO(3)\,,\; S = T\wedge \in so(3)\}$$

which has the structure of an algebraic variety in $\mathbf{RP}^8$ [10] as well as that of a differentiable manifold [31] (in fact it is exactly the tangent bundle of the rotation group, $TSO(3)$). We now show that for a slight modification of $\tilde{E}$ it is possible to find an explicit local coordinate homeomorphism.

Theorem 5.1 *Let $d_{\mathbf{x},\mathbf{x}'}(\mathbf{Q})$ be the triangulation function*[2], *which gives the depth of a point from its motion $\mathbf{Q}$ and its projective coordinates $\mathbf{x}$ at time t and $\mathbf{x}'$ at time $t+1$. Then $E \doteq \tilde{E} \cap d^{-1}_{\mathbf{x},\mathbf{x}'}(\mathbf{R}^2_+)$ is a topological manifold of class at least C_0.*

Proof:

E inherits the topology from $\mathbf{R}^9$. Consider the map

$$\begin{array}{rcl} \Phi : E & \rightarrow & \mathbf{S}^2 \times SO(3) \\ \mathbf{Q} & \mapsto & \begin{bmatrix} T \\ e^{\Omega\wedge} \end{bmatrix} = \begin{bmatrix} \pm\mathbf{V}_{.3} \\ \mathbf{U}R_Z(\pm\frac{\pi}{2})\mathbf{V}^T \end{bmatrix} \end{array} \quad (7)$$

where $\mathbf{U}, \mathbf{V}$ are defined by the Singular Value Decomposition (SVD) of $\mathbf{Q} = \mathbf{U}\Sigma\mathbf{V}^T$, $\mathbf{V}_{.3}$ denotes the third column of $\mathbf{V}$ and $R_Z(\frac{\pi}{2})$ is a rotation of $\frac{\pi}{2}$ about the Z axis. As usual Ω is the rotation 3-vector corresponding to the 3×3 rotation matrix $\mathbf{U}R_Z(\frac{\pi}{2})\mathbf{V}^T$ and is obtained using the Rodrigues' formulæ [25]. T is represented in spherical coordinates. Note that the map Φ defines the local coordinates of the essential manifold modulo a sign in the direction of translation and in the rotation angle of R_Z, therefore the map Φ associates to each element of the essential manifold 4 distinct points in local coordinates.

[2]See equation (8) in the proof for an instance of realization of the triangulation function.

This ambiguity can be resolved by imposing the *"positive depth constraint"*, i.e. each visible point lies in front of the observer [20, 21]. Consider one of the four local counterparts of $\mathbf{Q} \in E$, and the function $d_{\mathbf{X},\mathbf{X}'} : E \rightarrow \mathbf{R}^{1+1}$ defined by

$$d_{\mathbf{X},\mathbf{X}'}(\mathbf{Q}) = [Z, Z']^T \quad (8)$$

with $Z = \frac{<\mathbf{n}^i, \mathbf{m}^i>}{\|\mathbf{m}^i\|^2} \quad \forall i = 1, \ldots, n, \quad \mathbf{m}^i = (R\mathbf{x}^i) \wedge \mathbf{x}'^i$ and $\mathbf{n}^i = (RT) \wedge \mathbf{x}'^i$, which yields the depth of each point as a function of the projection and the motion parameters. Note that it is locally smooth away from zero translation. Now redefine the essential space as

$$\begin{array}{rcl} E & \doteq & \tilde{E} \cap d^{-1}_{\mathbf{X},\mathbf{X}'}(\mathbf{R}^2_+)^n = \\ & = & \{\mathbf{Q} = RS | R \in SO(3), S \in so(3), \|\mathbf{Q}\| = 1, \\ & & d_{\mathbf{X}^i,\mathbf{X}'^i}(\mathbf{Q}) > 0\, \forall i\} \end{array} \quad (9)$$

where $\mathbf{R}_+$ is the positive open real half space of $\mathbf{R}$, $d^{-1}_{\mathbf{X},\mathbf{X}'}$ denotes the preimage of $d_{\mathbf{X},\mathbf{X}'}$. Consider Φ restricted to E. It follows from the properties of the SVD that Φ is continuous, and furthermore it is bijective. It can be shown [9] that $\mathbf{Q} \in E \Leftrightarrow \Sigma = \mathrm{diag}\{1\ 1\ 0\}$ and hence the basis elements of the subspaces $< \mathbf{V}_{.1}, \mathbf{V}_{.2} >$ and $< \mathbf{U}_{.1}, \mathbf{U}_{.2} >$ are allowed to switch order. This happens, however, without affecting continuity of T and Ω. The inverse map is simply $\Phi^{-1} = e^{(\Omega\wedge)}(T\wedge)$ which is smooth. Hence E is a topological manifold of class at least C_0. ∎

5.2. Observability for $N \geq 8$ points

Since the essential constraint is linear in $\mathbf{Q}$, it is possible to write it using the notation

$$\chi(\mathbf{x}, \mathbf{x}')\mathbf{Q} = 0$$

where χ is a $n \times 9$ matrix and $\mathbf{Q}$ is interpreted as a nine-dimensional vector obtained by stacking the columns of $\mathbf{Q}$ on top of each other. The generic row of χ is $[xx',\ yx',\ x',\ xy',\ yy',\ y',\ x,\ y,\ 1\ ]$. Following the track of the previous sections, we will assume small acceleration or a statistical model for motion which, lifted to the essential manifold, results in a statistical model for $\mathbf{Q}$. The resulting model has the form

$$\begin{cases} \dot{\mathbf{Q}} = f(\mathbf{Q}, \nu_Q) & \mathbf{Q} \in E \\ \chi(\mathbf{x}, \mathbf{x}')\mathbf{Q} = \nu_\chi \end{cases}$$

where f is either a dynamical model or a statistical model; ν_Q and ν_χ are noise processes which can be characterized, as discussed in [28]. In [29], two recursive schemes are proposed for solving the estimation problem: one is based upon an Implicit Extended Kalman Filter (IEKF) in the local coordinates of the essential manifold, the other is based upon a linear update on the linear embedding space $\mathbf{R}^9$, followed by a projection onto the essential manifold.

Now consider χ: *if it has rank 8*, then there exists a unique $\mathbf{Q}$ which spans its null space modulo a sign, since we have imposed a constraint on its norm. This generates four distinct points in the local coordinates which reduce to a single solution once the positive depth constraint is imposed. Once this is done at one step, we choose a branch of the local coordinates map and stick with it for the subsequent time steps [10, 20, 32]. We are naturally led to the following:

Definition 5.1 *We say that the points* $\mathbf{x}$ *are in* general position $\stackrel{\cdot}{\Leftrightarrow}$ $rank(\chi(\mathbf{x}, \mathbf{x}')) = 8$.

Claim 5.1 *If an essential model is in general position then it is possible to reconstruct the motion* (V, Ω) *of the viewer modulo four solutions. The solution is unique once the positive depth constraint is imposed at one time instant.*

We still have to address the issue of the conditions under which the matrix χ has full rank. Furthermore, we need to deal with the case of less than 8 visible points, since it automatically excludes general position conditions.

5.3. Observability with less than 8 points

When less than 8 points are visible, it is not possible to achieve the above sufficient conditions for motion observability. Suppose, at time $t + \tau_i$, the matrix $\chi(t + \tau_i)$ has a null space of dimension k_i. *When the viewer moves with small acceleration*, we may write

$$\begin{array}{rcl} \chi(t)\mathbf{Q}(t) & = & 0 \\ \chi(t+\tau_1)\mathbf{Q}(t+\tau_1) = \chi(t+\tau_1)\mathbf{Q}(t) & = & 0 \\ \vdots & = & \vdots \\ \chi(t+\tau_p)\mathbf{Q}(t+\tau_p) = \chi(t+\tau_p)\mathbf{Q}(t) & = & 0 \end{array}$$

until $k_0 + k_1 + \ldots + k_p = 8$. If this happens, we can restate the sufficient conditions for motion observability for the extended matrix

$$\bar{\chi}_p \doteq \begin{bmatrix} \chi(t) \\ \chi(t+\tau_1) \\ \vdots \\ \chi(t+\tau_p) \end{bmatrix}.$$

Definition 5.2 *We say an essential model is in* general position *(GP) when either there are more than 8 visible points and* χ *has rank 8, or there exists a time instant* τ_p *such that the extended matrix* $\bar{\chi}_p$ *has rank 8.*

5.4. General position: rank condition for global observability of rigid motion

We are now interested in writing explicitly the general positions condition. This is done using results in [10, 21, 32] for the case of more than 8 points. The claim, extended to our general position condition, may be stated as:

Theorem 5.2 *An essential model is in general position* $\Longleftrightarrow$ *there does not exist a (proper) quadric surface in* $\mathbb{R}^3$ *which contains all the visible points and the path of the center of projection.*

Remark 5.1 *We report here a proof given by Mennucci* [24] *for the case of more than 8 visible points. The original proof by Longuet-Higgins may be found in* [21]. *The general case is obtained by substituting* $\mathbf{x}_{t+\tau_i}$ *for* $\mathbf{x}_i$. *Note that the quadric surface is a thin set in the 3D Euclidean space. The measurement noise in the projected coordinates is sufficient to set the model in general position. Note also that* $T \neq 0$ *plays a critical role in achieving global observability, while* Ω *(or* R*) has no influence.*

Remark 5.2 *There are situations in which the model* is not *in general position, for instance when observing one single point while holding constant velocity (the center of projection describes an arc of a circle, and we could fit a quadric passing through the observed point). The noise in the data will indeed set the model in general position; however, questions of conditioning arise when close to a singular (non-general position) configuration.*

Proof:

Let $T \neq 0$. Consider the points to be fixed in an intermediate reference system, where their coordinates are $(\mathbf{X}_i'')$ such that $\mathbf{X}_i' = R(\mathbf{X}_i'' - T), \mathbf{X}_i = R^T(\mathbf{X}_i'' + T)$; then $\mathbf{X}_i'^T \mathbf{Q} \mathbf{X}_i = 0 \;\; 1 \leq i \leq n$, and the same holds for $\mathbf{x}$ in place of $\mathbf{X}$. By substitution we get

$$\begin{aligned} \mathbf{x}_i'^T \mathbf{Q} \mathbf{x}_i &= [R(\mathbf{x}_i'' - T)]^T \mathbf{Q} R^T (\mathbf{x}_i'' + T) = \\ &= (\mathbf{x}_i'' - T)^T R^T \mathbf{Q} R^T (\mathbf{x}_i'' + T) = 0 \;\; \forall i. \end{aligned} \quad (10)$$

We may change the variable in this equation to be $\mathbf{Q}' = R^T \mathbf{Q} R^T$; since R is invertible, this would not change any of the considerations below. We will therefore assume $R = I$ without loss of generality.

$$\begin{aligned} &(\mathbf{x}_i'' - T)\mathbf{Q}(\mathbf{x}_i'' + T) = \\ &= \mathbf{x}_i''^T \mathbf{Q} \mathbf{x}_i'' - T^T \mathbf{Q} \mathbf{x}_i'' + \mathbf{x}_i''^T \mathbf{Q} T - T^T \mathbf{Q} T = 0 \;\; \forall i \end{aligned} \quad (11)$$

Call $< \mathbf{Q} > \doteq \{\mathbf{Q} \in \mathbb{R}^{3\times 3} \mid (\mathbf{x}_i'' - T)^T \mathbf{Q} (\mathbf{x}_i'' + T) = 0, \;\; 1 \leq i \leq n\}$; $< \mathbf{Q} >$ is a vector subspace of $\mathbb{R}^{3\times 3}$, and the fact that there is only one solution is equivalent to saying that the dimension of $< \mathbf{Q} >$ is one; indeed, $\dim(< \mathbf{Q} >)$ is always bigger or equal than one, since it contains the matrix $T\wedge$, as can be seen by direct substitution in eq. (11). Suppose that the equation (10) holds for a matrix M, and decompose it in the symmetric and antisymmetric part $A = \frac{M - M^T}{2}, S = \frac{M + M^T}{2}$, then

$$\mathbf{x}_i''^T S \mathbf{x}_i'' - 2T^T A \mathbf{x}_i'' - T^T S T = 0 \;\; 1 \leq i \leq n.$$

Consider the set $< V > \doteq \{x \in \mathbf{R}^3 \mid x^T S x - 2T^T A x - T^T S T = 0\}$. This set always contains the two points T and $-T$, the centers of projection (as a simple computation shows). Suppose there is no (proper) quadric surface containing the points $\mathbf{x}_i''$; then it must be that $V = \mathbf{R}^3$, that means that $S = 0$ and $T^T A = 0$; this means that M is necessarily a multiple of $T\wedge = \mathbf{Q}$, so we get that $\dim(< \mathbf{Q} >) = 1$.

Vice versa, suppose that the symmetric part S of M is nonzero or that $T^T A \neq 0$; then the set $< V >$ is a quadric surface that contains the points $\mathbf{x}_i''$ (by definition), and it contains the points T and $-T$, which are the two centers of projection (if the symmetric part $S = 0$, then the set $\{x \in \mathbf{R}^3 \mid T^T A x = 0\}$ is a plane, which is in any case a quadric surface). ■

6. Conclusions

We have studied the observability of rigid motion under projection. The model which defines the problem for feature points in the Euclidean 3D space lacks local observability. The observable manifold is covered with three levels of Lie differentiation. The problem is indeed observable, once formulated in the appropriate topological space.

We have then studied a formulation of visual motion estimation in terms of identification of an implicit dynamical model with parameters on the essential manifold [29]. The model is globally observable/identifiable with zero level of bracketing. When more points are available, redundancy may be exploited to reduce the effect of measurement noise.

Acknowledgements

The author wishes to thank Prof. Pietro Perona for his advice, directions, interaction and support. Prof. Ruggero Frezza and Prof. Giorgio Picci were very helpful with discussions and suggestions. The discussions with Prof. Richard Murray, Michiel van Nieuwstadt and Andrea Mennucci were crucial in understanding the geometric structure of the problem. This research has been funded by the California Institute of Technology, a fellowship from the "Fondazione Ing. Aldo Gini", a fellowship from the University of Padova–Italy, ONR grant N00014-93-1-0990, an AT&T Foundation Special Purpose grant. A long version of this paper has been submitted to the IFAC journal "Automatica".

References

[1] A. Azarbayejani, B. Horowitz, and A. Pentland. Recursive estimation of structure and motion using relative orientation constraints. *Proc. CVPR*, New York, 1993.

[2] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. RPL-TR 9107, Queen's University Kingston, Ontario, Robotics and perception laboratory, 1992. Also in Proc. CVPR 1992, pp 236-242.

[3] W. Boothby. *Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 1986.

[4] T. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. *IEEE trans. PAMI*, Jan. 1986.

[5] Bryant, Chern, Goldberg, and Goldsmith. *Exterior Differential Systems*. Mathematical Research Institute. Springer Verlag, 1992.

[6] R.S. Bucy. Non-linear filtering theory. *IEEE Trans. A.C. AC-10, 198*, 1965.

[7] W. Dayawansa, B. Ghosh, C. Martin, and X. Wang. A necessary and sufficient condition for the perspective observability problem. –, 1993.

[8] E. D. Dickmanns and V. Graefe. Dynamic monocular machine vision. *Machine Vision and Applications*, 1:223–240, 1988.

[9] O. Faugeras. *Three dimensional vision, a geometric viewpoint*. MIT Press, 1993.

[10] O. D. Faugeras and S. Maybank. Motion from point matches: multiplicity of solutions. *Int. J. of Computer Vision*, 1990.

[11] D.B. Gennery. Tracking known 3-dimensional object. In *Proc. AAAI 2nd Natl. Conf. Artif. Intell.*, pages 13–17, Pittsburg, PA, 1982.

[12] B. Ghosh, M Jankovic, and Y. Wu. Perspective problems in systems theory and its application in machine vision. *Journal of Math. Systems, Est. and Control*, 1994.

[13] D. Heeger and A. Jepson. Subspace methods for recovering rigid motion i: algorithm and implementation. *Int. J. Comp. Vision vol. 7 (2)*, 1992.

[14] J. Heel. Dynamic motion vision. *Proc. of the DARPA image understanding workshop*, 1989.

[15] C.C. Ho and N.H. McClamrock. Autonomous spacecraft docking using a computer vision systm. *Proc. of the 31st CDC – Tucson, AZ*, 1992.

[16] A. Isidori. *Nonlinear Control Systems*. Springer Verlag, 1989.

[17] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[18] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. of the ASME-Journal of basic engineering.*, 35-45, 1960.

[19] K. Hashimoto T. Kimoto T. Ebine H. Kimura. Manipulator control with image-based visual servo. *IEEE intl' Conference on Robotics and Automation, pages 2267–2272*, 1991.

[20] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[21] H.C. Longuet-Higgins. Configurations that defeat the eight-point algorithm. *Mental processes: studies in cognitive science, MIT press*, 1987.

[22] L. Matthies, R. Szeliski, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *Int. J. of computer vision*, 1989.

[23] S. Maybank. *Theory of reconstruction from image motion*. Springer Verlag, 1992.

[24] Andrea C. Mennucci. Recovering ego motion: an alternative proof for the configurations that defeat the eight-point longuet-higgins' algorithm. *Personal communication*, Jan. 1994.

[25] R.M. Murray, Z. Li, and S.S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

[26] H. Nijmeijer and A.J.Van Der Shaft. *Nonlinear Dynamical Control Systems*. Springer Verlag, 1990.

[27] J. Oliensis and J. Inigo-Thomas. Recursive multi-frame structure from motion incorporating motion error. *Proc. DARPA Image Understanding Workshop*, 1992.

[28] S. Soatto, R. Frezza, and P. Perona. Motion estimation via dynamic vision. *Submitted to the IEEE Trans. on Automatic Control. Registered as Technical Report CIT-CDS-94-004, California Institute of Technology. Reduced version in the proc. of the 33 IEEE Conference on Decision and Control. Available through the Worldwide Web Mosaic* (http://avalon.caltech.edu/cds/techreports/), 1994.

[29] S. Soatto, R. Frezza, and P. Perona. Motion estimation on the essential manifold. *In "Computer Vision ECCV 94, Lecture Notes in Computer Sciences vol. 801", Springer Verlag*, May 1994.

[30] S. Soatto and P. Perona. Dynamic visual motion estimation from subspace constraints. *Submitted to the Int. J. of Computer Vision. Registered as Technical Report CIT-CDS 94-005, California Institute of Technology. Reduced version in the proceeding of the first IEEE Int. Conf. on Image Processing*, 1994.

[31] S. Soatto and P. Perona. Structure-independent visual motion control on the essential manifold. *Technical Report CIT-CDS-94-013, California Institute of Technology. Short version in Proc. of the IFAC Symposium on Robot Control, Capri–Italy*, 1994.

[32] J. Weng, T. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(5):451–476, 1989.